ᗺ OPEN ACCESS

A Statistical Outlook into the Distribution of Crimes in Nigeria Using Principal Component Analysis

Guobadia Emwinloghosa Kenneth[1, 2*]

[1]Department of Administration, Federal Medical Centre, Asaba, Delta State, Nigeria
[2]Department of Statistics, University of Benin, Benin, Nigeria

*Corresponding author: Guobadia Emwinloghosa Kenneth

Abstract		Original Research Article

This paper analyses Nigeria crime data which consists of the averages of 50 crimes reported to the police for the period 2009 – 2019. The offenses include robbery, car theft, breakage of houses and stores, theft/stealing, serious accident and injury, murder, rape and assault, etc. Correlation analysis and main component analysis (PCA) were used to clarify the relationship between crimes and to assess the distribution of crimes in Nigeria. The result shows that on the average there are 14600.33 crimes in 2009, 14437.83 crimes in 2010, 14942.92 crimes in 2011, 12904.83 crimes in 2012, 12063.42 crime in 2013, 10342.58 crimes in 2014, 6677.667 crimes in 2015, 10602.75 crimes in 2016, 9357.75 crimes in 2017, 9093 crimes in 2018 and 14600.33 crimes in 2019. The result has shown that there is no significant correlation among Crimes against Persons which means that none of the variables can be used to predict or explain one another. The correlations among Crimes against Property are at least moderate. However, there are significant correlations among murder, armed robbery, grievous harm and wounding. There is also a significant correlation among assault, kidnapping, attempted murder, thief/ other stealing, store breaking, house breaking and escape from lawful arrest. The PCA has suggested retaining six components that explain about 95.24 percent of the total variability of the data set.
Keywords: Crime data, correlation analysis, principal component analysis, Nigeria, eigenvalues.

INTRODUCTION

Crime is a well-known word that is unique to every country, culture, and community in the world. Law and order is a given manner in which society regulates crimes by the nation(s). Crime is a threat that calls for close surveillance and study. Therefore, the pace at which crimes are committed in society and methods of curbing them need to be examined. Crime in society may not be eradicated, but it is necessary to have steps to curb crime [1]. Despite the fact that crime in society is unavoidable, different control and preventive steps have been taken and are still being taken to minimize the threat. Crime doesn't have a cap. It has little to do with developing, developing, or under-developing a world. In reality, in most of the world's developed countries, the worst crimes are published daily in the news. Therefore, crime differs across continents and its occurrence can be defined by the country. This implies that, for some countries, cultures and societies, some crimes are unique. In most industrialized countries, kidnapping, homicides, armed robbery and killing involving the use of armed arms are peculiar, for example. Suicide bombing, Rape, Homicides, Religion crime and Massacre are quite known in the Arab nations. Kidnapping, Ethnic fight, Tribal war, Corruption, Rioting, Religion clashes etc., are peculiar in Africa, etc.

Our study focuses only on Nigeria. In categorizing crimes, crimes do vary even within States of a nation. In Nigeria today, for example, terrorism, suicide bombing, religious war and killings, vandalism, etc., is a crime unique in the northern part of Nigeria, particularly in states such as Borno, Adamawa, Kano and Yobe, while in Nigeria, abduction, pipeline vandalism, armed robbery and oil theft are the order of the day. We have learnt of incidents of ritual killing, theft & stealing, kidnapping, attacks, wounding, armed robbery, etc. in the South West and in the Eastern part of Nigeria [2]. This means that the incidence of crime may vary across boundaries, but no crime-free culture, group or country exists. Therefore, there are 14 main

Citation: Guobadia Emwinloghosa Kenneth, A Statistical Outlook into the Distribution of Crimes in Nigeria Using Principal Component Analysis. Sch J Phys Math Stat, 2021 Jan 8(1): 1-7.

1

types of crimes associated with people in the Delta state. Theft & Other Stealing was classified by the police department as the state's most committed crime. Several causes of crime assistance in communities include, but are not limited to, illiteracy, poverty, broken families, poor industries, the environment, faith and extremism of religion, police inability and power to administer justice. The motivating factors often differ depending on the country, just as crimes vary across the globe. In most industrialized countries where the standard of living is considerably good enough, as a crime-motivating factor, one can overrule illiteracy and poverty. In country B, what motivates crime in country A may be less important. All these factors are substantially important to crime rates in most of our communities when examining our own situation for this report [3].

METHODOLOGY

In order to test the links between the crimes and the discrepancies between the offenses perpetrated in Nigeria between 2009 and 2019, and to decide which of the crimes is relevant to another. In order to understand the distributions in crime rates over the span of 10 years, we carried out some descriptive analysis of the results. In other words, to assess the rate of crime on an annual basis and to understand the prevalence of crime over crime over crime. The Principal Component Analysis enabled us to analyze the important components of crime rates in the state of Delta more thoroughly [4].

Principal Component Analysis

The Principal Component Analysis (PCA) is a statistical approach that uses an orthogonal transformation to translate a set of observations of potentially correlated variables into a set of values called principal components of linearly uncorrelated variables. The goal is to minimize a larger number of variables into a smaller set of 'artificial' variables, known as 'primary components,' which account for much of the original variables' variance. It means that the components extracted have a high degree of relationship with the original variables if these variables are strongly correlated. This allows you to understand if any of the variables you have selected are not appropriately descriptive of the construct you are interested in, and should be omitted from your new calculation. The number of key components is equal to or less than the number of initial variables. This transformation is described in such a way that the first main component has the greatest possible variance (that is, accounts for as much of the data variability as possible) and each subsequent component in turn has the greatest possible variance under the restriction that the previous components are orthogonal [5-7]. The resulting vectors are an orthogonal basis set that is uncorrelated. The key components are orthogonal since they are the covariance matrix's own vectors, which are symmetric. The relative scaling of the original variables is subject to the PCA. PCA is conceptually distinct from factor analysis, often used interchangeably with factor analysis in practice, and is included in SPSS within the 'Factor Method' [8-10]

METHOD

The key idea of the principal component (PC) transformation is to check for a few (< p) derived variables that retain most of the information provided by the variance of the p random variables. Let X be a vector of p random variables [1]. Let the random vector $X' = [X_1, X_2, ..., X_p]$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3, ... , \geq \lambda_p \geq 0$.

Considering the linear combinations,

$$Y_j = \alpha_j^| X = \alpha_{j1}^| X_1 = \alpha_{j2}^| X_2 + \cdots + \alpha_{jp}^| X_p = \sum \alpha_{jk}^| X_k$$

$j = 1,2,3, ..., p$ of the element of X, where α_j is a vector of p components $\alpha_{j1}^|, \alpha_{j2}^|, ..., \alpha_{jk}^|$

Then $Var (Y_j) = \alpha_j^| \sum \alpha_j$
And $Cov (Y_j) = \alpha_j^| \sum \alpha_k$
Where $j = 1,2, ... , p$

The PCs are the uncorrelated linear combinations $Y_1, Y_2, ..., Y_p$ whose variances are as large as possible [2]. We focus on the variances by finding the PCs. The first step is to search for a linear combination of $\alpha_1^| X$ with maximum variance, so that $\alpha_j^| X = \alpha_{j1}^| X_1 = \alpha_{j2}^| X_2 + \cdots + \alpha_{jp}^| X_p$

Then, search for a linear combination $\alpha'_2 X$ uncorrelated with $\alpha_1^| X$ having the maximum variance, and so on, so that a linear combination $\alpha_k^| X$ is discovered at the kth point that has the maximum variance uncorrelated with $\alpha_1^| X = \alpha_2^| X, ..., \alpha_{k-1}^| X$. The kth derived variable $\alpha_K^| X$ is the kth PC. It is possible to find up to p PCs, but we have to stop after the qth point ($q \leq p$) when q PCs have accounted for most of the variance in X.

The variance of a PC is equal to the PC's own eigenvalue,

$$Var (Yj) = \alpha_j^| \sum \alpha_j \text{ Where } j = 1,2, ... , p$$

The total variance in a data set is equal to the total variance of PCs

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum Var (X_i) = \lambda_1 + \lambda_2 + \lambda_3, ..., \geq \lambda_p = \sum Var (Y_i)$$

Using a common standardization method of transforming all data to have zero mean and unit standard deviation, the data was standardized for the variables to be of similar size. The corresponding standardized variables for a random vector $X' = [X_1, X_2, ..., X_p]$ are;

$$Z = \left(\frac{X_j - \mu_j}{\sigma_{11}}\right)$$

Denoted in matrix notation as;

$$Z = V^{1/2}(X - \mu).$$

Where $V^{1/2}$ is the diagonal standard deviation matrix. Thus $E(Z) = 0$ and $Cov(Z) = P$. The PCs of Z can be obtained from eigenvectors of the correlation matrix ρ of X. All our previous properties for X are applied for the Z, so that the notation Y_j refers to the jth PC and (λ_j, α_j) refers to the eigenvalue – eigenvector pair. However, the quantities derived from Σ are not the same from those derived from ρ [2].

The jth PC of the standardized variables $Z = [z_1, z_2, ..., z_p]$ with Cov (Z), is given by

$$Y_j = \alpha_j^| Z = \alpha_j^|(V^{1/2}) - 1(X - \mu)$$

So that,

$$\Sigma Var(Y_j) = \Sigma Var(Z_j) = p \ j = 1,2, ..., p$$

In this case, (λ_1, α_1), (λ_2, α_2), ..., (λ_p, α_p) are the eigenvalue- eigenvector pairs with $\lambda_1 \geq \lambda_2 \geq \lambda_3, ..., \geq \lambda_p > 0$

Interpretation of the Principal Components

The loading or the eigenvector $\alpha_j = \alpha_1, \alpha_2, \alpha_3, ..., \alpha_j$ is the calculation of the significance for a given PC of a calculated variable. The first element is a weighted average of the variables when all elements are positive, which is often referred to as a measure of the overall rate of crime. The positive and negative coefficients in the subsequent components may also be viewed as a form of component of crime [3, 4]. Visual interpretation is enhanced by the graph of the first two or three loadings against each other. The score is an indicator of a PC's significance for an observation. By simply substituting the original variables Xij into the first q PCs package, the new PC observations Y_{ij} are obtained. $Y_{ij} = \alpha_{j1}^| X_{i1} = \alpha_{j2}^| X_{i2} + \cdots + \alpha_{jp}^| X_{ip} \ j = 1, 2, 3 ...., p; \ i = 1, 2.., n.$ Visual understanding strengthens the plot of the first two or three PCs against each other [11].

The proportion of Variance

The variance proportion tells us the PC that described the original variables best. For deciding the number of components to be retained in the analysis, a total proportion of explained variance is a useful criterion. A Scree plot provides a strong graphical illustration of the PCs' ability to describe the data variations [12-13].

ANALYSIS AND RESULT

Table-1: Crime rates across the year (2009-2019)

Year	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Total Crime	175204	173254	179315	154858	144761	124111	80132	127233	112293	109116	175204

Table-2: Correlation Matrix

	Murder	Assault	Suicide	Rape	Kidnapping	Armed Robbery	Burglary	Forgery	Gambling
Murder	1.000	-.074	-.082	.374	.290	.885	-.394	-.091	-.155
Assault	-.074	1.000	.457	-.262	.602	-.295	.812	.533	.370
Suicide	-.082	0.457	1.000	.309	-.025	-.357	.420	.321	-.156
Rape	.374	0.262	.309	1.000	-.476	.394	-.464	.133	-.705
Kidnapping	.290	.602	-.025	-.476	1.000	.090	.481	.282	.601
Armed Robbery	.865	-.295	-.357	.394	.090	1.000	-.695	-.298	-.291
Burglary	-.394	.812	.420	-.464	.481	-.695	1.000	.646	.488
Forgery	-.091	.533	.321	.133	.282	-.298	.646	1.000	.041
Gambling	-.155	.370	-.156	-.705	.601	-.291	.488	.041	1.000
Breach of Public Peace	.473	.700	.238	-.353	.737	.201	.462	.230	.441
Perjury	.411	.201	.083	.255	.070	.100	.144	.039	-.037
Bribery and Corruption	-.392	.775	.113	-.357	.360	-.336	.563	.216	.178
Attempted Murder	.382	.824	.309	-.157	.660	.224	.463	.317	.153
Man Slaughter	-.718	-.525	.397	-.173	-.645	-.495	-.147	-.215	-.102
Grievous Harm and Wounding	.587	.546	.085	-.068	.755	.535	.088	.047	.259

Attempted Suicide	-.276	.500	-.152	-.809	.666	-.392	.618	.170	.953
Child Stealing	-.629	.684	.209	-.251	.133	-.561	.573	.262	.267
Slave dealing	.434	-.419	.042	.379	.071	.506	-.644	-.415	-.223
Thief/Other stealing	-.339	.944	.346	-.483	.564	-.493	.33	.528	.466
House breaking	-.244	.925	.324	-.550	.632	-.439	.896	.536	.534
Store breaking	-.299	.802	.274	-.672	.674	-.507	.894	.503	.670
Escape from Lawful Arrest	-.519	.847	.481	-.413	.473	-.643	.843	.438	.456

Table-2: Correlation Matrix Contd. I.

	Breach of Public Peace	Perjury	Bribery and Corruption	Attempted Murder	Man Slaughter	Grievous Harm and Wounding	Attempted Suicide	Child Stealing	Slave Dealing
Murder	.473	.411	-.392	.382	-.718	.587	-.276	-.629	.434
Assault	.700	.201	.775	.824	-.525	.546	.500	.684	-.419
Suicide	.238	.083	.113	.309	-.397	.085	-.152	.209	.042
Rape	-.353	.255	-.357	-.157	-.173	-.068	-.809	-.251	.379
Kidnapping	.737	.070	.360	.660	-.645	.755	.666	.133	.071
Armed Robbery	.201	.100	-.336	.224	-.495	.535	-.392	-.561	.506
Burglary	.462	.144	.563	.463	-.147	.088	.618	.573	-.644
Forgery	.230	.039	.216	.317	-.215	.047	.170	.262	-.415
Gambling	.441	-.037	.178	.153	-.102	.259	.953	.267	-.223
Breach of Public Peace	1.000	.190	.294	.859	-.784	.752	.488	.061	-.110
Perjury	.190	1.000	.004	.177	-.225	.055	-.139	-.066	-.178
Bribery and Corruption	.294	.004	1.000	.584	-.099	.346	.373	.890	-.415
Attempted Murder	.859	.177	.584	1.000	-.804	.827	.254	.296	-.174
Man Slaughter	-.784	-.225	-.099	-.804	1.000	-.887	-.068	.122	-.305
Grievous Harm and Wounding	.752	.055	.346	.827	-.887	1.000	.264	.087	.287
Attempted Suicide	.488	-.139	.373	.254	-.068	.264	1.000	.412	-.320
Child Stealing	.061	-.066	.890	.296	.122	.087	.412	1.000	-.440
Slave dealing	-.110	-.178	-.415	-.174	-.305	.287	-.320	-.440	1.000
Thief/Other stealing	.586	-.019	.824	.688	-.287	.378	.641	.769	-.540
House breaking	.691	-.021	.723	.722	-.353	.415	.695	.638	-.550
Store breaking	.647	-.147	.566	.581	-.283	.339	.808	.516	-.492
Escape from Lawful Arrest	.367	-.125	.779	.502	-.191	.273	.595	.819	-.385

Table-2: Correlation Matrix Condt II

	Thief/Other stealing	House Breaking	Store Breaking	Escape from Lawful Arrest
Murder	-.339	-.244	-.299	-.519
Assault	.944	.925	.802	.847
Suicide	.346	.324	.274	.481
Rape	-.483	-.550	-.672	-.413
Kidnapping	.564	.632	.671	.473
Armed Robbery	-.493	-.439	-.507	-.643
Burglary	.883	.896	.894	.843
Forgery	.528	.536	.503	.438
Gambling	.466	.534	.670	.456
Breach of Public Peace	.586	.691	.647	.367
Perjury				
Bribery and Corruption	-.019	-.021	-.147	-.125
Attempted Murder	.824	.723	.566	.779
Man Slaughter				
Grievous Harm and Wounding	.688	.722	.581	.502
Attempted Suicide	-.287	-.353	-.283	-.191
Child Stealing	.378	.415	.339	.273
Slave dealing				
Thief/Other stealing	.641	.695	.808	.595
House breaking	.769	.638	.516	.819
Store breaking	-.540	-.550	-.492	-.385
Escape from Lawful Arrest	1.000	.980	.909	.925
	.980	1.000	.958	.871
	.909	.958	1.000	.841
	.925	.871	.841	1.000

In short, there is a very clear positive link between murder and armed robbery. With Breach of Peace, assaults have a clear positive connection. Attempted Murder has a clear positive correlation with serious injury. Store Breaking has a good association with Attempted Suicide that is optimistic. With Man Slaughter, slave trafficking has a clear positive connection. The good or positive correlation means that the other also increases as one increases. For example, murder has a positive association with armed robbery, meaning that armed robbery cases increase as murder cases increase. This suggests that, certain murder cases may be as a result of rape, arson, armed robbery, suicide, grievous harms/wounding, attacks, shop breaking, house breaking, stealing, false pretense, illegal possession, breach of peace, forgery, abduction of children and unusual offences [14-15]. That is, all of these crimes are linked, so one could lead to another.
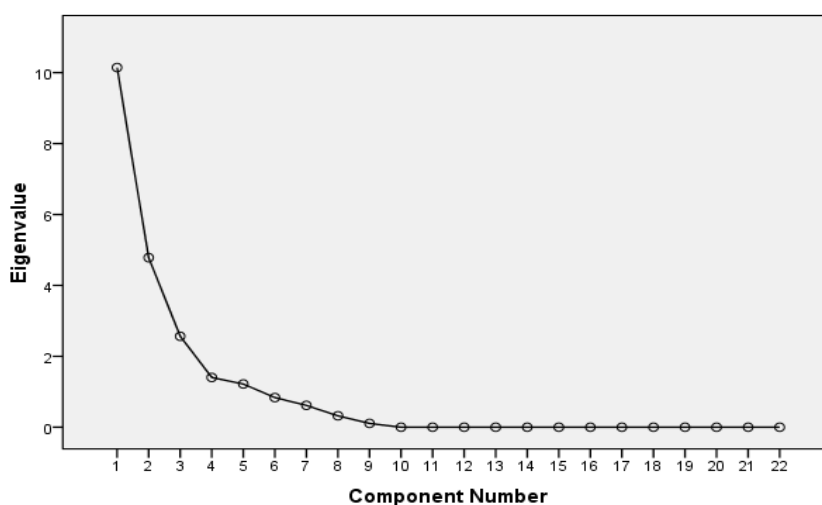

Fig-1: Screen Plot of the Principal Components

Table-3 displays the eigenvalues and the total proportions of the explained variance. It would be rational to maintain the first six PCs, given the eigenvalue-one criterion and the Scree plot in Figure-1. A widely accepted rule states that holding only PCs with values greater than 1. is appropriate. It may, however, be seen from table 3 that only the first six pcs can be retained to describe 95.24 percent of the total variability.

Table-3: Eigenvalues

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of variance	Cumulative %
1	10.145	46.113	46.113	10.145	46.113	46.113
2	4.782	21.736	67.849	4.782	21.736	67.849
3	2.566	11.663	79.511	2.566	11.663	79.511
4	1.404	6.383	85.895	1.404	6.383	85.895
5	1.219	5.542	91.437	1.219	5.542	91.437
6	.837	3.803	95.239			
7	.617	2.803	98.042			
8	.322	1.466	99.508			
9	.108	.492	100.000			
10	1.141E-15	5.185E-15	100.000			
11	6.716E-16	3.053E-15	100.000			
12	4.288E-16	1.949E-15	100.000			
13	3.787E-16	1.721E-15	100.000			
14	2.447E-16	1.112E-15	100.000			
15	1.335E-16	6.066E-16	100.000			
16	3.936E-17	1.789E-16	100.000			
17	5.974E-18	2.715E-17	100.000			
18	-9.110E-17	-4.141E-16	100.000			
19	-1.929E-16	-8.769E-16	100.000			
20	-2.932E-16	-1.333E-15	100.000			
21	-3.138E-16	-1.426E-15	100.000			
22	-6.295E-16	-2.861E-15	100.000			

The above Table-3 centered on the six PCs that describe the retention of 95.24 percent of the total variability of the data collection. Component 1 has a positive (weak and strong) relationship with all reported crimes, but it primarily recognizes (strong relationship with) murder, grievous harm/wounding, assaults, stealing/theft, rape/indecent abuse, robbery, false pretense/cheating, arson, forgery, breach of peace and unnatural crimes as the key crimes committed or concentrated crimes. Component 2 describes the key offenses committed as forgery, illegal possession, shop breaking and stealing/theft. It has a negative relationship with Forgery and Unlawful Ownership (decrease) and a positive relationship with Shop Breaking and Stealing/Theft (increase). Burglary/Stealing is only known as the centered offense in Component 3. This suggests that the daily crime in this area is burglary/stealing. With Suicide, component 4 increases and decreases with House breaking. If the other offense becomes history or decreases, we will use this position as a measure of new common crime. Component 5 determines the prevalence of arson and child stealing and Component 6 acknowledges no particular crime.

CONCLUSION

The results deduced from the paper are the following. Crimes against persons have no meaningful correlations, which means that none of the variables can be used to forecast (explain) each other. There is at least a mild link between crimes against land. There are, however, important similarities between murder, armed robbery, grievous damage and injury. There is also a strong link between assault, abduction, attempted murder, burglary, shop breaking, house breaking, and escape from legal detention.

The retained six PCs explain 95.24 percent of the total variability of the data set are retained. Component 1 has a positive relationship (weak and strong) with all the crimes recorded but majorly it identifies (strong relationship with) Murder, Grievous harm/wounding, Assaults, Stealing/Theft, Rape/Indecent Assault, Kidnapping, False pretense/Cheating, Arson, Forgery, Breach of Peace and Unnatural offences as the major crimes committed or concentrated offences. Component 2 identify Forgery, Unlawful possession, Store Breaking and Stealing/Theft as the major crimes committed. It has a negative relationship (decrease) with Forgery and Unlawful possession and a positive relationship (increase) with Store breaking and Stealing/Theft.

Component 3 identify Burglary/Stealing only as the concentrated offence. This implies that, Burglary/Stealing is the daily offence in this location. Component 4 increases with Suicide and decreases with House breaking. We can view this location as measure of new common offence while the other offence is becoming history or reducing. Component 5 identify Arson and Child Stealing and Component 6 identify no specific crime as prevalence.

These components helped us in identifying the distribution of crimes committed in Nigeria. Thus, identifying the distribution of these crimes in Nigeria will allow the investors, tourist and travelers to measure the level of risk and to plan preventive measures for safeguarding their investments, property and lives.

REFERENCES

1. Obubu Maxwell, Ikediuwa Udoka Chinedu, Anabike Charles Ifeanyi, and Nwokike Chukwudike Chinonso (2019). On Modeling Murder Crimes in Nigeria. *Scientific Review*. 5(8), 157-162.
2. Osuji, G.A., Okoro, C.N., Obubu, M., Obiora-Ilouno H.O. (2016). Effect of Akaike Information Criterion on Model Selection in Analyzing Auto-Crash Variables. *International Journal of Sciences: Basic and Applied Research*. 26(1):98-109.
3. Cattell, R.B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research*, 1,245-276.
4. Durkheim, E. (1933). The division of Labour in Society. *Macmillan, New York*.
5. Gabor, T. (1978). Crime displacement: the literature and strategies for its investigation. *Crime and Justice*. 6:100 - 7.
6. Gulumbe et al (2008). Analysis of Crime Data using Principal Component Analysis. *International Journal of Academic Research in Business and Social Sciences*. 2(6).
7. Obubu Maxwell, Babalola A. Mayowa, Ikediuwa U. Chinedu, Amadi Peace (2018). Modeling Count Data; a Generalized Linear Model Framework. American Journal of Mathematics and Statistics. 8(6): 179-183 doi:10.5923/j.ajms.20180806.03.
8. Louis, S., Cookie, W. S., Louis, A. Z. and Sheldon, R. E. (1981). Human Response to Social Problems. *The Dorsey Press*, Illinois.
9. Robin Beaumont; Principal Component Analysis & Factor Analysis: Using SPSS 19 and R (psych package)
10. Maritz, J. (2010). Honest answers to your questions about investing in Nigeria: Will I have to fear for my safety in Nigeria? June, 14, 2010.http://www.tradeinvest Nigeria. com/news/623915.htm
11. Rencher, A.C. (2002) Methods of Multivariate Analysis. 2nd edition, John Wiley & Son, New York.
12. Richard, A.J. and Dean, W.W. (2001). Applied Multivariate Statistical Analysis. 3rd edition, Prentice-Hall, New Delhi.
13. Jolliffe, I.T. (2002). Principal Component Analysis. 2nd edition, Springer-Verlag, New York. CBN Journal of Applied Statistics. 3(2), 49-60.
14. Obubu, M., Konwe C.S., Nwabenu D.C., Omokri Peter A., Chijoke M. (2016) Empirical Analysis on Road Traffic Crashes in Anambra State, Nigeria; Accident Prediction Modeling using Regression Approach. European Journal of Statistics and Probability. 4(4):18-27
15. Osuji G.A., Obubu, M., Obiora-Ilouno H.O (2015). An Investigation on Crime Rate in Southeastern Nigeria. European Journal of Statistics and Probability. 3(4):1-9.